Research Article

Image Reconstruction Using Analysis Model Prior

Yu Han,[1] Huiqian Du,[1] Fan Lam,[2] Wenbo Mei,[1] and Liping Fang[3]

[1]*School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*
[2]*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[3]*School of Mathematics, Beijing Institute of Technology, Beijing 100081, China*

Correspondence should be addressed to Huiqian Du; duhuiqian@bit.edu.cn

The analysis model has been previously exploited as an alternative to the classical sparse synthesis model for designing image reconstruction methods. Applying a suitable analysis operator on the image of interest yields a cosparse outcome which enables us to reconstruct the image from undersampled data. In this work, we introduce additional prior in the analysis context and theoretically study the uniqueness issues in terms of analysis operators in general position and the specific 2D finite difference operator. We establish bounds on the minimum measurement numbers which are lower than those in cases without using analysis model prior. Based on the idea of iterative cosupport detection (ICD), we develop a novel image reconstruction model and an effective algorithm, achieving significantly better reconstruction performance. Simulation results on synthetic and practical magnetic resonance (MR) images are also shown to illustrate our theoretical claims.

1. Introduction

Sparse sampling theory [1–3] plays a key role in a broad spectrum of techniques involved in signal and image processing over the past decade. It states that an unknown signal can be recovered from a small number of random linear measurements given that the signal is sparse. Consider a sparse signal or image, vectorized as $\mathbf{x} \in \mathbb{R}^n$, which has very few nonzero elements in the sense that $\|\mathbf{x}\|_0 \ll n$, where $\|\mathbf{x}\|_0$ is the count of the nonzeros in $\mathbf{x}$. We expect to reconstruct $\mathbf{x}$ by solving the following l_0-norm minimization problem,

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \quad \|\mathbf{x}\|_0$$
$$\text{subject to} \quad \mathbf{y} = \mathbf{Mx}, \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{m \times n}$ ($m < n$) denotes the measurement matrix that produces the measurement vector $\mathbf{y} \in \mathbb{R}^m$. However, this is an NP-hard problem which prompts us to look for alternatives to solve it in an approximate fashion. A popular and effective way is to rewrite (1) as the basis pursuit problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \quad \|\mathbf{x}\|_1$$
$$\text{subject to} \quad \mathbf{y} = \mathbf{Mx}. \tag{2}$$

As a matter of fact, this is not the general case since signals or images do not exhibit sparsity directly but have sparse representations in specific transform domains. Broadly speaking, there are two data models for describing signals. The first one is the sparse synthesis model [4, 5], wherein $\mathbf{x}$ is assumed to admit sparse representation in a fixed dictionary $\mathbf{D} = \{\mathbf{d}_i\}_{i=1}^d \in \mathbb{R}^{n \times d}$. Put differently, $\mathbf{x} = \mathbf{Dc}$ can be viewed as a linear combination of very few dominant atoms from $\mathbf{D}$. Thus, the reconstruction process (2) is reformulated as

$$\hat{\mathbf{c}} = \underset{\mathbf{c}}{\operatorname{argmin}} \quad \|\mathbf{c}\|_1$$
$$\text{subject to} \quad \mathbf{y} = \mathbf{MDc}; \tag{3}$$
$$\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{c}}.$$

As we know, a tremendous surge of effort has been devoted to studying the sparse synthesis model, and much progress has been made ranging from theoretical foundations [6, 7] to appealing applications, including denoising [8], inpainting [9], and more. Additionally, a long series of algorithms [10–13] together with provable guarantees [14, 15] have been put forward.

While the synthesis model has gained widespread attention, a similar alternative was proposed modeling signals from an analysis perspective [16–20]. Mathematically, given an analysis operator $\boldsymbol{\Omega} = \{\boldsymbol{\omega}_j\}_{j=1}^p \in \mathbb{R}^{p \times n}$, a signal belonging to the analysis model is supposed to admit a sufficiently sparse analysis representation $\mathbf{z} = \boldsymbol{\Omega}\mathbf{x} \in \mathbb{R}^p$. In particular, we are interested in cases where $p \geq n$, so that various information within $\mathbf{x}$ can be captured [17]. The commonly used analysis operators include the finite difference [21], overcomplete wavelet transforms [22], and the curvelet transform [23]. Just in contrast to the "sparsity" in the synthesis setting, the analysis model concentrates on the zero-valued coefficients. A fundamental notion measuring the quantity of the zeros is defined as cosparsity

$$l = p - \|\boldsymbol{\Omega}\mathbf{x}\|_0. \tag{4}$$

The l-cosparse analyzed vector $\mathbf{z}$ is assumed to have l unknown locations of the zeros, referred to as the cosupport

$$\Lambda = \left\{ j \mid \left\langle \boldsymbol{\omega}_j, \mathbf{x} \right\rangle = 0, \ j \in \{1, \dots, p\} \right\}. \tag{5}$$

Recent studies have theoretically shown that the analysis model has its own advantage over the synthesis one [17]. Moreover, adopting the analysis model leads to a collection of successful applications, such as denoising [23, 24], inpainting [25], and medical imaging [26, 27]. Models mentioned above simply exploit the (co)sparsity prior that is implicit in signals or images. However, (co)sparsity alone is insufficient for making reasonable inferences from these models, and the minimum measurement requirement for reconstruction is improved limitedly. Obviously, the trade-off between (co)sparsity and measurements can be further improved by imposing *a priori* knowledge. A common way to incorporate prior knowledge is through the use of the signal support, which has been extensively studied in the synthesis context [28–30]. However, relatively little research has been devoted to imposing prior knowledge in the analysis sense. In this work, we wish to investigate the consequences of exploiting analysis model prior. We concentrate on image reconstruction given the *a priori* cosupport knowledge. First, we formulate the reconstruction problem in terms of the analysis operator $\boldsymbol{\Omega}$ in general position, which means that every set of n rows from $\boldsymbol{\Omega}$ are linearly independent [17]. In such a case, we derive the minimum number of measurements for uniquely determining a cosparse solution to the inverse problem $\mathbf{y} = \mathbf{M}\mathbf{x}$. The resulting number is smaller than that of the standard analysis model without using prior cosupport. Second, we dive into details of the model associated with the 2D finite difference operator whose rows show significant linear dependencies [17]. We also provide an improved minimum measurement requirement for guaranteeing the uniqueness given the prior cosupport. Having a theoretical foundation for the analysis uniqueness properties, we develop a novel image reconstruction method based on the idea of iterative cosupport detection (ICD). This two-stage scheme proceeds by alternatively calling its two key components: image reconstruction and cosupport detection. At each iteration, the image is reconstructed using the cosupport knowledge extracted from the previous iteration. After the acquiring of the image estimate, one can identify even better cosupport to be used in the next iteration. Moreover, when performing the cosupport detection, a multidirectional finite difference is considered. Therefore, the detection and the reconstruction parts work together enabling us to gradually obtain reliable cosupport and a reasonable image estimate. Simulation results on synthetic and practical magnetic resonance (MR) images demonstrate the effectiveness and show considerable improvement of the proposed method compared with other regularization based methods. This consequently indicates that, through the use of the analysis model prior, we can achieve a given reconstruction quality with fewer measurements or alternatively obtain a better reconstruction under the same measurement requirement.

The remainder of this paper is organized as follows. In Section 2, a detailed description of our proposed method is given from two aspects: analysis operators in general position and the 2D finite difference operator. Uniqueness issues are also explored. Section 3 describes the methods used for comparison and rules for image quality assessment. Section 4 presents simulation results and discussion which validate our theoretical claims. Finally in Section 5, conclusions are summarized.

2. Method

Analysis model prior has been successfully used for many signal processing tasks but has been done with little theoretical justification. In this section, we focus on the analysis-based reconstruction given that the cosupport is known *a priori*. The number of measurements required for guaranteeing the unique reconstruction is proved to be essentially reduced. We first consider the analysis operators in general position.

2.1. Analysis Operators in General Position

2.1.1. The Cosupport Is Exactly Known. Consider l-cosparse image $\mathbf{x} \in \mathbb{R}^n$ whose cosupport is Λ and a redundant analysis operator $\boldsymbol{\Omega} \in \mathbb{R}^{p \times n}$ $(p \geq n)$ in general position; namely, every set of n rows from $\{\boldsymbol{\omega}_j\}_{j=1}^p$ are linearly independent [17]. The cosparse representation analysis model where $\mathbf{x}$ belongs is related to the recently proposed union-of-subspaces model [31–33]:

$$\mathbf{x} \in \bigcup_{|\Lambda|=l} \mathscr{W}_\Lambda, \tag{6}$$

which is a union of all the $\binom{p}{l}$ possible subspaces of dimension $n-l$. Here, $\mathscr{W}_\Lambda = \text{Null}(\boldsymbol{\Omega}_\Lambda)$ denotes the null-space of the analysis operator $\boldsymbol{\Omega}$ indexed by Λ. $|\Lambda|$ is the cardinality of Λ. The rows associated with Λ, $\{\boldsymbol{\omega}_j\}_{j\in\Lambda}$, define the analysis subspace. Remark that removing rows from $\boldsymbol{\Omega}$ for which $\langle \boldsymbol{\omega}_j, \mathbf{x} \rangle \neq 0$ leaves the subspace unchanged.

Consider the aforementioned noiseless linear inverse problem:

$$\mathbf{y} = \mathbf{M}\mathbf{x}, \tag{7}$$

where $\mathbf{M} \in \mathbb{R}^{m \times n}$ and Ω are assumed to be mutually independent. Provided that the cosupport Λ is exactly known, the cosparse $\mathbf{x}$ is supposed to satisfy the linear system:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{M} \\ \Omega_\Lambda \end{bmatrix} \mathbf{x}. \tag{8}$$

The fact that (8) identifies a unique $\mathbf{x}$ is equivalent to the requirement:

$$\mathscr{W}_\Lambda \cap \mathrm{Null}\,(\mathbf{M}) = \{\mathbf{0}\}. \tag{9}$$

This requirement indicates that the minimum number of measurements is:

$$m \geq \max_\Lambda \dim\,(\mathscr{W}_\Lambda) = n - l. \tag{10}$$

2.1.2. The Cosupport Is Unknown. Actually, the case we do care about is that we only know the cosparsity level l, whereas the cosupport Λ is undetermined. The uniqueness issue in this sense has been explored in [17].

Lemma 1. *Let $\bigcup_\Lambda \mathscr{W}_\Lambda$, $|\Lambda| = l$, be a union of l-cosparse analysis subspaces induced by the analysis operator Ω. Then the linear system $\mathbf{y} = \mathbf{M}\mathbf{x}$ admits a unique l-cosparse solution if and only if for any $|\Lambda_1|, |\Lambda_2| = l$*

$$\mathscr{W}_{\Lambda_1, \Lambda_2} \cap \mathrm{Null}\,(\mathbf{M}) = \{\mathbf{0}\}, \tag{11}$$

where

$$\begin{aligned} \mathscr{W}_{\Lambda_1, \Lambda_2} &= \mathscr{W}_{\Lambda_1} \oplus \mathscr{W}_{\Lambda_2} \\ &= \{\mathbf{u} \mid \mathbf{u} = \mathbf{x}_1 + \mathbf{x}_2,\ \mathbf{x}_1 \in \mathscr{W}_{\Lambda_1},\ \mathbf{x}_2 \in \mathscr{W}_{\Lambda_2}\}. \end{aligned} \tag{12}$$

Clearly, for any $\mathbf{u} \in \mathscr{W}_{\Lambda_1, \Lambda_2}$, we have $\mathrm{cosupp}(\mathbf{u}) \supseteq \Lambda_1 \cap \Lambda_2$ and $|\Lambda_1 \cap \Lambda_2| \geq 2l - p$. Therefore, we can conclude that the minimum number in terms of the cosparsity level l is

$$m \geq \max_{\Lambda_1, \Lambda_2} \dim\,(\mathscr{W}_{\Lambda_1, \Lambda_2}) = n + p - 2l. \tag{13}$$

2.1.3. The Cosupport Is Inaccurately Known. More generally, we are considerably interested in the case where the cosupport is inaccurately known. In other words, a superset of the true cosupport is available, which comes up in many applications. Even in the absence of available prior knowledge, one can still extract useful information from the current solution and use it subsequently.

We still assume $\mathbf{x} \in \mathbb{R}^n$ is l-cosparse with cosupport Λ. Λ_0 is the prior cosupport containing small errors denoted by Λ'. Then, the true cosupport of $\mathbf{x}$ can be expressed as

$$\Lambda = \Lambda_0 \setminus \Lambda'. \tag{14}$$

We state the condition which enables us to guarantee the uniqueness of the linear inverse problem (7), namely, the minimum measurement requirement.

Proposition 2. *Assume that $\mathbf{x} \in \bigcup_\Lambda \mathscr{W}_\Lambda$, $|\Lambda| = l$, is the l-cosparse image to be reconstructed. Λ_0 is the prior cosupport*

with small error $\Lambda' \subseteq \Lambda_0$. Given $|\Lambda_0| = l_0 < p$ and $|\Lambda'| = l' \ll l_0$, the minimum number of measurements for identifying a unique solution to the linear system $\mathbf{y} = \mathbf{M}\mathbf{x}$ is

$$m \geq n + l_0 - 2l. \tag{15}$$

Proof. As we have seen, the prior cosupport Λ_0 is not exactly consistent with the true cosupport Λ of $\mathbf{x}$; that is, there exists small error Λ' in Λ. It means that we are supposed to consider the invertibility of $\mathbf{M}$ over the direct sum $\mathscr{W}_{\Lambda_1, \Lambda_2}$ of any two subspaces $\mathscr{W}_{\Lambda_1}, \mathscr{W}_{\Lambda_2} \subseteq \bigcup_\Lambda \mathscr{W}_\Lambda$. Assume any two l-cosparse images $\mathbf{x}_1 \in \mathscr{W}_{\Lambda_1}$, $\mathbf{x}_2 \in \mathscr{W}_{\Lambda_2}$ with corresponding cosupports $\Lambda_1 = \Lambda_0 \setminus \Lambda_1'$ and $\Lambda_2 = \Lambda_0 \setminus \Lambda_2'$, respectively. Note that $|\Lambda_1| = |\Lambda_2| = l$ and let $|\Lambda_1'| = |\Lambda_2'| = l'$. Then, we have $l = l_0 - l'$. Consider any $\mathbf{u}$ from $\mathscr{W}_{\Lambda_1, \Lambda_2}$ as defined in (12). The cosupport of $\mathbf{u}$ is obtained as

$$\begin{aligned} \mathrm{cosupp}\,(\mathbf{u}) &\supseteq \Lambda_1 \cap \Lambda_2 = (\Lambda_0 \setminus \Lambda_1') \cap (\Lambda_0 \setminus \Lambda_2') \\ &= (\Lambda_0 \cap (\Lambda_1')^c) \cap (\Lambda_0 \cap (\Lambda_2')^c) \\ &= \Lambda_0 \cap (\Lambda_1' \cup \Lambda_2')^c = \Lambda_0 \setminus (\Lambda_1' \cup \Lambda_2'), \end{aligned} \tag{16}$$

which yields

$$|\mathrm{cosupp}\,(\mathbf{u})| \geq l_0 - 2l'. \tag{17}$$

Thus, we have

$$\dim\,(\mathscr{W}_{\Lambda_1, \Lambda_2}) \leq n - (l_0 - 2l') = n + l_0 - 2l. \tag{18}$$

Consequently, we conclude that the minimum number of measurements

$$m \geq \max_{\Lambda_1, \Lambda_2} \dim\,(\mathscr{W}_{\Lambda_1, \Lambda_2}) = n + l_0 - 2l, \tag{19}$$

which completes the proof. $\qquad\square$

Proposition 2 states that the measurement number required for uniquely identifying the solution to (7) tends to be smaller in the sense that the cosupport is known *a priori*. As l_0 approaches p, m is equivalent to that of the case without prior knowledge; namely, $n + p - 2l$. When the prior cosupport selected is consistent with the ground-truth, the minimum measurement number is achieved as $n - l$.

2.2. Specific 2D Finite Difference Analysis Operator. In this subsection, we would like to investigate the family of finite difference operators on graphs, Ω_{DIF} (significantly related to total variation norm minimization [21]), which has been proven successful for sparse signal and image recovery [34–36]. We explore uniqueness issues in the sense that cosupport knowledge is known *a priori*. Due to the fact that this class of analysis operators exhibit strong linear dependencies [17], the theoretical results derived above cannot be applied directly. Our analysis is based upon the work [17]. For ease of notation, we will drop the subscript DIF and simply use Ω to represent Ω_{DIF} hereafter when there is no ambiguity. Assume that Ω is defined on a 2D nonoriented graph with vertices V and edges E. An edge $e = (v_1, v_2)$ connecting two vertices v_1 and v_2 can

be regarded as a finite difference. Λ is a subset of E and the set of vertices connected by at least one edge in Λ is denoted by $V(\Lambda)$ which is composed of $J(\Lambda)$ connected components. A connected component is a collection of vertices connected to one another by a walk through vertices within Λ. Thus, the dimension of $\mathcal{W}_\Lambda = \text{Null}(\boldsymbol{\Omega}_\Lambda)$ is given by

$$\dim\left(\mathcal{W}_\Lambda\right) = |V| - |V(\Lambda)| + J(\Lambda), \tag{20}$$

where $|V| - |V(\Lambda)|$ denotes the number of isolated vertices which have distinct values from all the neighbors. Suppose that the known part of the cosupport is denoted by Λ_0 and the number of the connected components is $J(\Lambda_0) := J$. Then, in terms of the cosparsity level l, we present a concrete bound for $\dim(\mathcal{W}_\Lambda)$, which reveals the minimum measurement number.

Proposition 3. *Let $\boldsymbol{\Omega}$ be the 2D finite difference analysis operator that computes horizontal and vertical discrete derivatives of an $n = N \times N$ image $\mathbf{x}$. The measurement matrix $\mathbf{M}$ is assumed to be mutually independent from $\boldsymbol{\Omega}$. For a fixed l and the known cosupport Λ_0 which corresponds to J connected components, the equation $\mathbf{y} = \mathbf{Mx}$ admits at most one solution with cosparsity l only if*

$$m \geq 2n - l - \sqrt{2Jl} + J \geq 2\max_\Lambda \dim\left(\mathcal{W}_\Lambda\right). \tag{21}$$

Proposition 3 reveals that the measurements required for uniquely determining the cosparse solution to the inverse problem can be reduced given that some part of the cosupport is known. Additionally, the minimum measurement number decreases monotonically as the number of the connected components J increases. When $J = 1$, the minimum number of measurements $m \geq 2n - l - \sqrt{2l} + 1$, which is equivalent to the case that no cosupport information is available. For completeness, the proof can be found in Appendix.

2.3. Proposed Reconstruction Model and Algorithm

2.3.1. Reconstruction Model. Armed with the above-described theoretical analyses, we are now in position to solve the reconstruction problem regularized with cosupport prior. More specifically, we expect to constrain the cosparsity of $\mathbf{x}$ within the prior cosupport Λ_0:

$$\begin{aligned}\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \quad & \left\|(\boldsymbol{\Omega}\mathbf{x})_{\Lambda_0}\right\|_0 \\ \text{subject to} \quad & \mathbf{y} = \mathbf{Mx}.\end{aligned} \tag{22}$$

However, as previously mentioned, the l_0-norm involved in the combinatorial minimization program is an NP-hard penalty and thus is not feasible to be solved for practical applications. An effective and widely used alternative is the l_1-relaxation:

$$\begin{aligned}\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \quad & \left\|(\boldsymbol{\Omega}\mathbf{x})_{\Lambda_0}\right\|_1 \\ \text{subject to} \quad & \mathbf{y} = \mathbf{Mx},\end{aligned} \tag{23}$$

in which the l_1-norm enables promoting high cosparsity in the solution. Its desirable convexity facilitates various computationally tractable algorithms, and much recent progress in the theory of analysis l_1-minimization has been made [37, 38]. In this work, we focus on the analysis l_1-recovery which is given in an unconstrained fashion. The objective function is formulated as a linear combination of the data consistency error and a modified cosparsity-inducing penalty:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \left\|\mathbf{y} - \mathbf{Mx}\right\|_2^2 + \lambda \mathcal{T}_\Lambda(\mathbf{x}), \tag{24}$$

where $\lambda > 0$ controls the influence between the fidelity term and the regularization term:

$$\mathcal{T}_\Lambda(\mathbf{x}) = \sum_{i=1}^{4} \left\|(\boldsymbol{\Omega}_i\mathbf{x})_{\Lambda_i}\right\|_1. \tag{25}$$

Remark that the analysis operator we employed in our method is the 2D finite difference operator involving four directional components; that is, $\{\boldsymbol{\Omega}_i\}_{i=1}^{4}$, respectively, compute discrete derivatives in vertical, horizontal, and two diagonal directions. $\{\Lambda_i\}_{i=1}^{4}$ denote the associated four-direction cosupport sets that are detected iteratively.

2.3.2. Algorithm. Having presented the reconstruction model, we now turn to the question how to effectively solve it. We propose a two-stage algorithm based on the idea of iterative cosupport detection (ICD). The proposed ICD allows extracting the reliable information of the underlying solution and enables achieving a reasonable reconstruction. This two-stage scheme proceeds by alternatingly calling its two components: image reconstruction and cosupport detection. In the reconstruction step, we solve a truncated l_1-minimization problem via conjugate gradient method using the cosupport knowledge obtained from the previous iteration. ICD will terminate if the approximate solution is accurate enough. Otherwise, cosupport detection will be performed in light of this inaccurate reconstruction, thereby yielding a better $\{\Lambda_i\}_{i=1}^{4}$ to be used in the next iteration. Consequently, the detection and the reconstruction parts of the proposed ICD work together enabling us to gradually obtain reliable cosupport and a reasonable image estimate.

It is important to note that the proposed ICD requires reliable cosupport detection. However in most cases, it is hard to completely avoid false detections at each iteration. To this end, we address ways to identify the index sets $\{\Lambda_i\}_{i=1}^{4}$. Actually, the analyzed vectors in each direction $\{\boldsymbol{\Omega}_i\mathbf{x}\}_{i=1}^{4}$ are not strictly cosparse but exhibit a strong decay. A conceptually simple detection strategy is to obtain the cosupport in a truncated manner. Without loss of generality, we assume $\{I_i\}_{i=1}^{4}$ as the four-direction index sets of $\{\boldsymbol{\Omega}_i\mathbf{x}\}_{i=1}^{4}$ after sorting in an ascending order. Then, the cosupport set in the ith direction is created as

$$\Lambda_i = \left\{I_i(k)\right\}_{k=1}^{L}, \tag{26}$$

where the truncated length L is fixed and specified in advance. Empirically, a relatively stable reconstruction can be achieved

provided that L is suitably scaled in a certain range. This strategy is easy to implement but has one drawback; that is, the assumption that the true cosupport is included in the inaccurate prior cosupport cannot be well satisfied when L is fixed. To alleviate this drawback, we propose another detection strategy which is more adaptive and effective. The cosupport sets $\{\Lambda_i\}_{i=1}^{4}$ are iteratively learned based on thresholding [28]:

$$\Lambda_i = \left\{ j \mid \left|(\Omega_i \mathbf{x})_j\right| < \beta_i \right\}, \tag{27}$$

where $(\Omega_i \mathbf{x})_j$ represents the jth element of $\Omega_i \mathbf{x}$. The threshold β_i is set as

$$\beta_i = \frac{\|\Omega_i \mathbf{x}\|_\infty}{\eta_i}, \quad (\eta_i > 0). \tag{28}$$

Here, η_i is the threshold parameter selected as an exponential function of the iteration number t:

$$\eta_i^t = w^{t-1}, \tag{29}$$

where $w > 1$ is a positive integer. For $t = 1$, we have $\beta_i^1 = \|\Omega_i \mathbf{x}\|_\infty$, indicating that no prior cosupport knowledge is exploited in the first iteration. In the sequel, the threshold decreases with the increase of the iteration so that the cosupport size reduces gradually. However, the cosupport in each direction is not strictly decreasing over the iteration, which allows the current detection to include indices within the true cosupport that are excluded from the detected cosupport in previous iterations. This leads to an attractive self-corrected capacity of the cosupport detection. The proposed algorithm, referred to as ICD, is outlined as follows.

Algorithm 4 (ICD). Consider the following:

(1) *Input.* Consider undersampled measurements $\mathbf{y}$, cosupport detection parameter w (or L), regularization parameter λ, and maximum iteration number $t_{\max}$.

(2) *Iteration.* While the stopping criterion is not reached, do the following:

  (i) *Image Reconstruction.* Solve the minimization problem (24) for the image estimate $\hat{\mathbf{x}}^t$ using the cosupport information of the previous iteration $\{\Lambda_i^{t-1}\}_{i=1}^{4}$.

  (ii) *Cosupport Detection.* Update the cosupport information $\{\Lambda_i\}_{i=1}^{4}$ using the detection criterion equation (27) (or (26)) based on $\hat{\mathbf{x}}^t$.

  (iii) Consider $t = t + 1$.

(3) *Output.* The reconstructed image $\hat{\mathbf{x}} = \mathbf{x}^t$.

3. Evaluation

To evaluate the performance of the proposed method, we performed simulations on both synthetic and practical MR images. Similar to prior work on CS-MRI [34, 39], we simulated the data acquisition by randomly sampling the 2D discrete Fourier transform coefficients of test images according to the patterns. Thus, the measurement matrix in (24) was defined as the undersampled Fourier transform $\mathbf{F_u}$. The number of compressive measurements was measured in terms of the percentage of total number of Fourier coefficients, namely, the sampling ratio (SR). All simulations were carried out under MatLab R2011b running on a PC with a 3.2 GHz processor and 4 GB memory. To assess the effectiveness of the proposed method, we compared it with other potential reconstruction techniques, including (i) SparseMRI [39]: the leading MR image reconstruction approach combining the wavelets and the standard total variation (TV) regularization, regardless of the effect of the wavelets, this approach can be approximately viewed as the proposed method without using cosupport information; thus the contribution of the integration of analysis *a priori* knowledge can be demonstrated; (ii) NLTV [40]: the well-known method based on wavelet sparsity and nonlocal TV penalty; (iii) ISD-TV [28–30]: the method constraining the sparsity of wavelet coefficients over the complement of the known support, which can be regarded as incorporating prior knowledge to the sparse synthesis model; in order to make a fair comparison, we added a TV term; (iv) SDBS-TV [41]: another synthesis-based method combining support knowledge and block-sparse property in the wavelet domain. The proposed method was tested using both truncated and threshold-based strategies, respectively, named as ICD-TR and ICD-TH. We also expected further enhancement by combining our method with the wavelet penalty, named as ICD-WT, which in general solves

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{F_u}\mathbf{x}\|_2^2 + \eta_1 \mathcal{T}_\Lambda (\mathbf{x}) + \eta_2 \|\Psi\mathbf{x}\|_1, \tag{30}$$

where Ψ denotes the wavelet transform operator and $\eta_1, \eta_2 > 0$. Remark that threshold-based strategy is used for cosupport detection in ICD-WT.

For quantitative evaluation, the reconstruction quality was measured by the relative l_2-norm error (RLNE) which is a standard image quality metric indicating the difference between the reconstruction $\mathbf{x}_r$ and the ground-truth $\mathbf{x}_g$

$$\zeta_{\mathrm{RLNE}} = \frac{\|\mathbf{x}_r - \mathbf{x}_g\|_2}{\|\mathbf{x}_g\|_2}. \tag{31}$$

As for practical MR images containing more fine features, the quality of the reconstruction is quantified by two other metrics. The first one is the high-frequency error norm (HFEN) [42], defined as

$$\zeta_{\mathrm{HFEN}} = \frac{\|\mathrm{LoG}(\mathbf{x}_r) - \mathrm{LoG}(\mathbf{x}_g)\|_2}{\|\mathrm{LoG}(\mathbf{x}_g)\|_2}, \tag{32}$$

where LoG is a rotationally symmetric Laplacian of Gaussian filter capturing detailed textures. The filter kernel is of size 15×15 pixels and with a standard deviation of 1.5 pixels, the same as that in [42]. The second one is the structural similarity (SSIM) index [43], comparing local patterns of

TABLE 1: RLNE results of test reconstruction methods with random undersampling scheme.

Test image	Sampling ratio	Method						
		SparseMRI	NLTV	ISD-TV	SDBS-TV	ICD-TR	ICD-TH	ICD-WT
Shepp Logan	RL = 10	0.2618	0.1451	0.2338	0.1267	0.0517	0.0390	0.0258
	RL = 11	0.2014	0.1196	0.1852	0.0892	0.0221	0.0117	0.0109
	RL = 12	0.1494	0.0873	0.1009	0.0522	0.0098	0.0042	0.0037
Brain-1	SR = 0.1	0.1773	0.1529	0.1540	0.1480	0.1362	0.1213	0.1131
	SR = 0.2	0.1075	0.0801	0.0803	0.0681	0.0653	0.0601	0.0549
	SR = 0.3	0.0615	0.0385	0.0426	0.0371	0.0309	0.0267	0.0257
	SR = 0.4	0.0346	0.0183	0.0269	0.0178	0.0151	0.0142	0.0139
	SR = 0.5	0.0224	0.0054	0.0184	0.0085	0.0052	0.0047	0.0045
Brain-2	SR = 0.1	0.2036	0.2057	0.1746	0.1681	0.1554	0.1477	0.1436
	SR = 0.2	0.1396	0.1154	0.1016	0.0990	0.0932	0.0875	0.0834
	SR = 0.3	0.0870	0.0710	0.0677	0.0583	0.0550	0.0514	0.0491
	SR = 0.4	0.0587	0.0368	0.0447	0.0411	0.0352	0.0313	0.0302
	SR = 0.5	0.0400	0.0302	0.0321	0.0249	0.0223	0.0206	0.0198

pixel intensities between $\mathbf{x}_r$ and $\mathbf{x}_g$ that have been normalized for luminance and contrast:

$$\zeta_{\text{SSIM}} = \sum_i \frac{\left(2\mu_{r,i}\mu_{g,i} + C_1\right)\left(2\sigma_{rg,i} + C_2\right)}{\left(\mu_{r,i}^2 + \mu_{g,i}^2 + C_1\right)\left(\sigma_{r,i}^2 + \sigma_{g,i}^2 + C_2\right)}. \tag{33}$$

Here, $\mu_{r,i}$ and $\mu_{g,i}$ are mean intensities at the ith local window of $\mathbf{x}_r$ and $\mathbf{x}_g$, while $\sigma_{r,i}$ and $\sigma_{g,i}$ are the corresponding standard deviations. $\sigma_{rg,i}$ denotes the covariance and the constants C_1, C_2 are included to avoid instability.

4. Results and Discussion

4.1. Shepp Logan Phantom. We first tested our method on an ideal example: the Shepp Logan numerical phantom of size 256×256. The undersampling scheme we employed is the 2D radial trajectory. Figure 1 shows the reconstruction results of different methods using 12 radial lines (RL). Reconstruction parameters in (24) were set as follows: $\lambda = 5 \times 10^{-4}$ and $L = 64000$ for ICD-TR and $w = 2$ for ICD-TH. We also tested the proposed method using 10 and 11 radial lines. The RLNE comparison results are presented in Table 1. The superiority of the proposed method is clearly seen from reconstructed images and error maps which were obtained by subtracting the reconstructions from the original image (shown at the same scale). The results indicate that methods without using any prior knowledge, such as SparseMRI and NLTV, cause serious artifacts. Methods imposing prior support knowledge in the synthesis sense, including ISD-TV and SDBS-TV, improve the reconstruction quality to a certain extent, but the edges are not well preserved. The proposed ICD significantly improves the reconstruction and is capable of suppressing more artifacts and preserving more details. As for different cosupport detection strategies, ICD-TH performs better than ICD-TR since the threshold-based strategy enables correcting the cosupport adaptively. Table 2 shows the number of true and false detections of four directional cosupports at each iteration using ICD-TH. It indicates that the number of false

detections in each direction decreases gradually until the true cosupport is detected. However, the number of true detections is not necessarily always increasing.

4.2. Practical MR Images. While encouraged by the results on the phantom image, we conducted further simulations on more realistic images to evaluate the practical effectiveness of the proposed method. The CS measurements were generated by undersampling the Fourier coefficients of the fully sampled (ground-truth) MR images whose intensities were normalized to a maximum magnitude of 1. The original T1-weighted image (Brain-1) of size 256×256 (courtesy of Professor N. Schuff at the UCSF School of Medicine) and the sampling mask of 30% sampling ratio are shown in Figure 2. The undersampling scheme we used is the variable density random pattern which is widely used in k_y-k_z plane for 3D imaging enabling removing the aliasing interference without degrading the image quality [39]. Reconstruction parameters were set as $\lambda = 5 \times 10^{-4}$, $L = 58000$ for ICD-TR, and $w = 5$ for ICD-TH. Figure 2 gives the reconstructed images and error maps through different methods, and Table 3 presents the HFEN and SSIM results. The capabilities of our method were also demonstrated under different sampling ratios as presented in Table 1. The reconstruction results show considerable improvement of the proposed method with respect to the subjective visual quality and quantitative indices. It has been noted from the error maps and the HFEN and SSIM results that the image reconstructed by our method has more fine features and is the closest to the ground-truth. The superior performance of ICD can be attributed to the use of the analysis model prior, namely, the cosupport knowledge. Additionally, we observe that the proposed ICD combining the wavelet penalty, ICD-WT, performs slightly better than ICD-TR and ICD-TH. However, by adding an extra penalty term, the algorithm is slowed down, which weakens the superiority of ICD-WT.

We also tested the proposed ICD on a T2-weighted image (Brain-2) which was acquired from the Siemens scanner 3.0T,
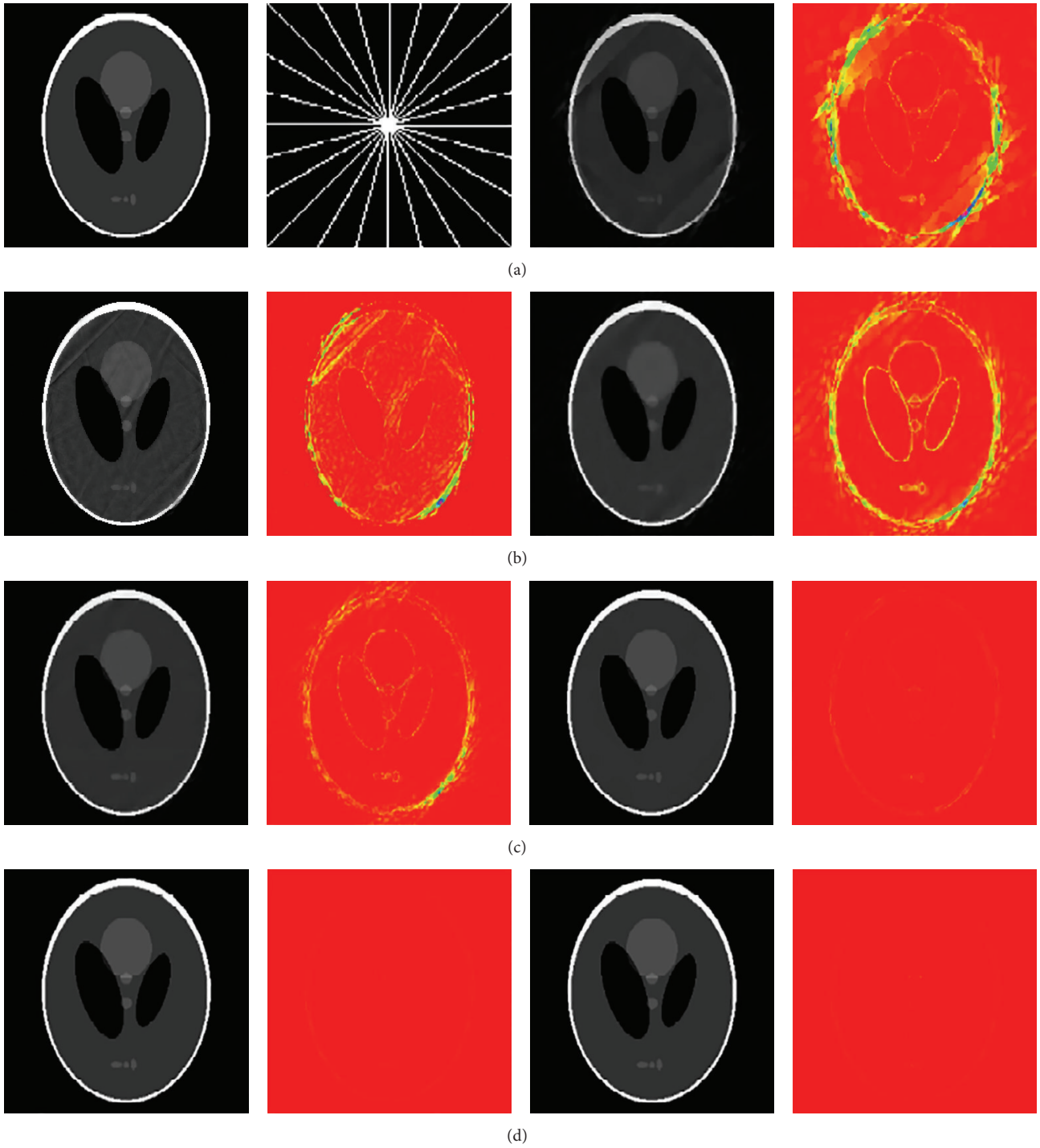
FIGURE 1: (a) (left to right) shows the original Shepp Logan phantom, the sampling mask, and the reconstructed and error images by SparseMRI. (b) shows the reconstructions by NLTV (left) and ISD-TV (right). (c) shows the reconstructions by SDBS-TV (left) and ICD-TR (right). (d) shows the reconstructions by ICD-TH (left) and ICD-WT (right).

SE sequence with imaging parameters: TR = 4000 ms, TE = 91 ms, slice thickness = 5.0 mm, flip angle = 120°, and the field of view (FOV) = 176 × 220 mm × mm. Reconstruction parameters were set as $\lambda = 5 \times 10^{-4}$, $L = 54000$, and $w = 5$. Figure 3 presents the reconstructed images and the error maps of different methods under 30% sampling ratio, while Table 3 gives the HFEN and the SSIM values. RLNE results under different sampling percentages are also given in Table 1. It is straightforward to see that the proposed method performs the best under all sampling ratios.

TABLE 2: The number of true/false detections of four directional cosupports over iterations.

Direction	Iteration	Cosupport size	True detections	False detections
Horizontal $\mathbf{\Omega}_1\mathbf{x}$	1	65074	64036	1038
	2	64601	64022	579
	3	64272	64053	219
	4	64055	64050	5
	5	64054	64054	0
Vertical $\mathbf{\Omega}_2\mathbf{x}$	1	65103	64472	631
	2	64849	64471	378
	3	64679	64472	207
	4	64473	64472	1
	5	64472	64472	0
Diagonal $\mathbf{\Omega}_3\mathbf{x}$	1	64993	63705	1288
	2	64374	63691	683
	3	64021	63719	302
	4	63721	63716	5
	5	63720	63720	0
Diagonal $\mathbf{\Omega}_4\mathbf{x}$	1	64810	63689	1121
	2	64371	63678	693
	3	63998	63707	291
	4	63706	63703	4
	5	63707	63707	0

TABLE 3: HFEN and SSIM results of practical MR images reconstructed using different methods.

Test image	Metric	Method						
		Sparse MRI	NLTV	ISD-TV	SDBS-TV	ICD-TR	ICD-TH	ICD-WT
Brain-1	ζ_{HFEN}	0.1628	0.0902	0.0971	0.0856	0.0628	0.0514	0.0480
	ζ_{SSIM}	0.9406	0.9740	0.9703	0.9731	0.9783	0.9851	0.9855
Brain-2	ζ_{HFEN}	0.1860	0.1547	0.1418	0.1198	0.1034	0.0946	0.0895
	ζ_{SSIM}	0.8908	0.9061	0.9138	0.9288	0.9357	0.9411	0.9443

4.3. Parameter Evaluation. In this subsection, we explore effects of the parameters involved in the proposed method. We begin by considering the regularization parameter λ, since the selection of the optimal λ-value is necessary. The reconstructions of Brain-1 and Brain-2 by ICD-TH are performed for different λ-values and under different sampling ratios. Motivated by the similar decay rates of four directional transform coefficients, we employed the same λ in each direction. Figure 4 displays the curves of RLNE values as a function of λ. The selected λ-values are marked with Stars. From the curves, the optimal λ-values under different sampling ratios are almost identically selected between $10^{-4} \sim 10^{-3}$.

Then, we consider effects of the cosupport detection parameters on the performance of the proposed method. As for ICD-TR, we are supposed to evaluate the truncated length L. Table 4 presents the RLNE results of both Brain-1 and Brain-2 under 30% sampling ratio with varying L. It can be observed that the proper range of L is image-dependent. However, the error undulates slightly when L ranges from 52000 to 60000, corresponding to 80~90% of the image dimension. The iterative process makes ICD-TR less sensitive to a few errors present in the cosupport. However with a fixed L, the true cosupport is not easy to be identified. We then consider ICD-TH which is more effective and adaptive. The selection of the threshold parameter w is essential in that it affects the convergence behavior of the reconstruction. The RLNE between the true and the reconstructed images at each iteration was computed and plotted for different values of $w = 2, 5, 8, 11$. From the error-iteration curves shown in Figure 5, we see that a large w leads to fast convergence but results in a relatively poor reconstruction. This is due to the fact that a large w makes the threshold decrease fast so that a number of indices belonging to the true cosupport are excluded from the detected cosupport. On the other hand, a small w allows a slowly changing threshold so that the cosupport can be corrected gradually. However, the convergence rate will not be satisfactory provided that w is too small. Based on our simulation results, we set $w = 2 \sim 5$ to achieve a trade-off between the convergence rate and the reconstruction quality.

The above-mentioned parameter evaluation is based on simulations on a large number of test images. The optimal ranges of the parameters we provided yield good empirical results. However, we realize that the parameters were chosen manually to minimize the reconstruction error. In practical applications, the ground-truth is not available. In these cases, one can conduct parameter selection using the effective
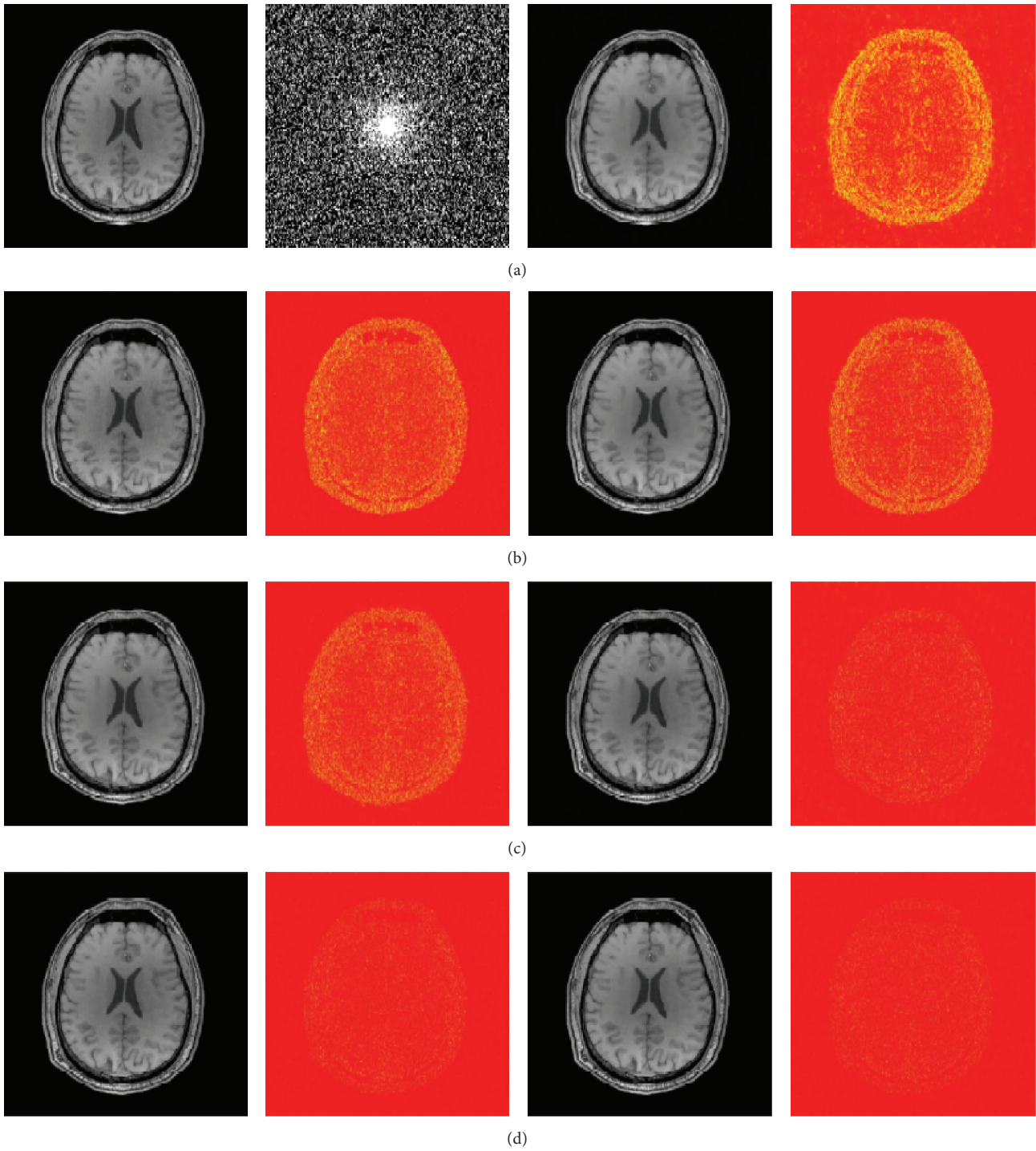
FIGURE 2: (a) (left to right) shows the original image (Brain-1), the sampling mask, and the reconstructed and error images by SparseMRI. (b) shows the reconstructions by NLTV (left) and ISD-TV (right). (c) shows the reconstructions by SDBS-TV (left) and ICD-TR (right). (d) shows the reconstructions by ICD-TH (left) and ICD-WT (right).

TABLE 4: RLNEs with varying truncated length L in ICD-TR reconstruction.

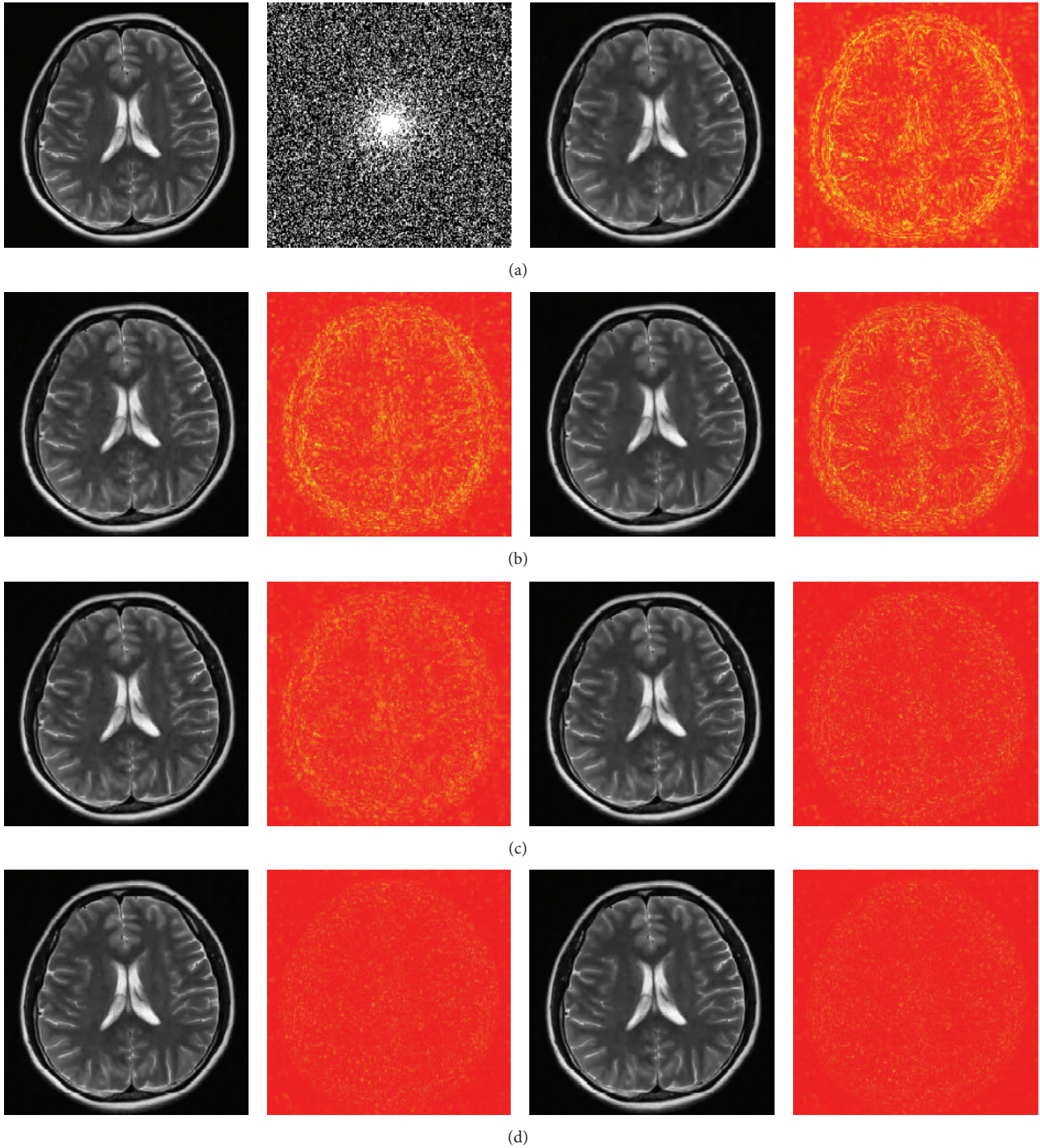| Test image	Truncated length $L(\times 10^3)$							
	50	52	54	56	58	60	62	64
Brain-1	0.0413	0.0384	0.0355	0.0320	0.0309	0.0336	0.0374	0.0443
Brain-2	0.0633	0.0575	0.0550	0.0563	0.0583	0.0592	0.0612	0.0651

(a)

(b)

(c)

(d)

FIGURE 3: (a) (left to right) shows the original image (Brain-2), the sampling mask, and the reconstructed and error images by SparseMRI. (b) shows the reconstructions by NLTV (left) and ISD-TV (right). (c) shows the reconstructions by SDBS-TV (left) and ICD-TR (right). (d) shows the reconstructions by ICD-TH (left) and ICD-WT (right).

L-curve strategy [44] or more sophisticated approaches [45, 46]. The discussion of these approaches is beyond the scope of this paper.

We also tested our method (ICD-TH) on Brain-1 and Brain-2 using radial and Cartesian patterns of 20% and 30% sampling ratios. From the results in Table 5, we see that

different patterns may affect the reconstruction quality. However, once the pattern is fixed, our method yields the best performance.

4.4. Comparison with Two-Direction Case. In this subsection, we conducted simulations on the Shepp Logan phantom with
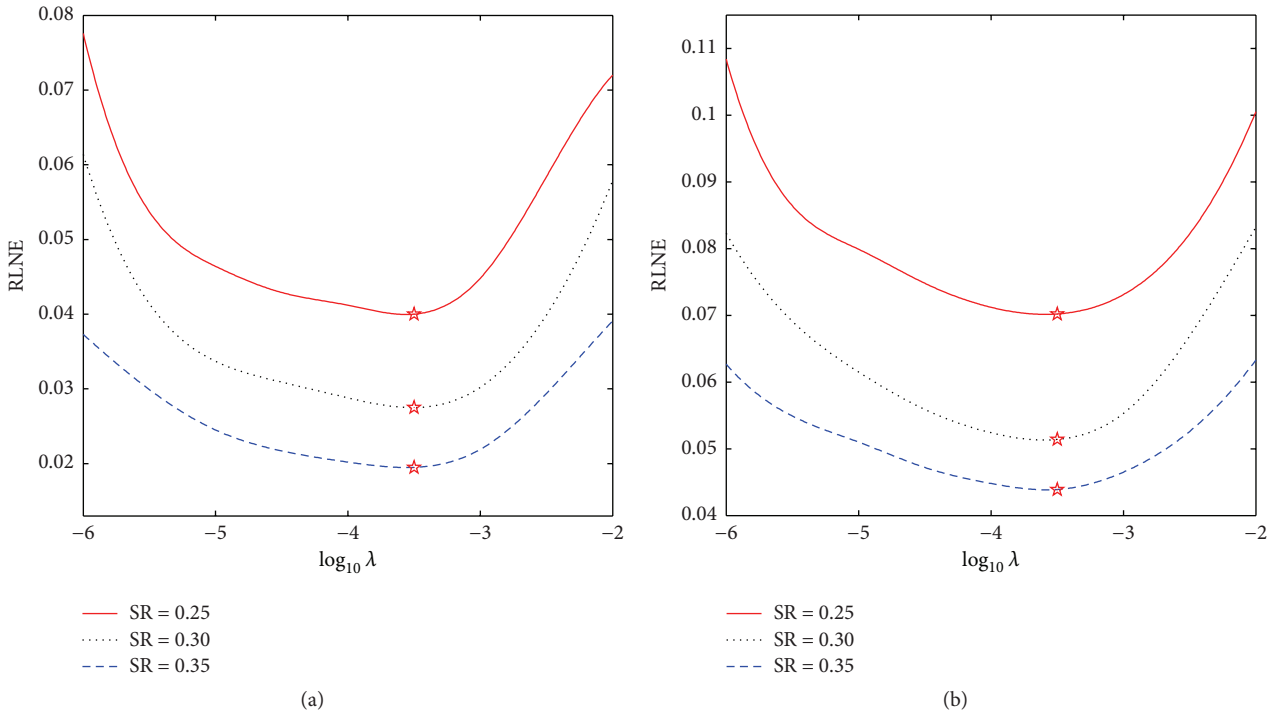
FIGURE 4: RLNEs versus regularization parameter λ for the reconstructions of Brain-1 (a) and Brain-2 (b) under different sampling ratios. Stars indicate the selected λ-values.
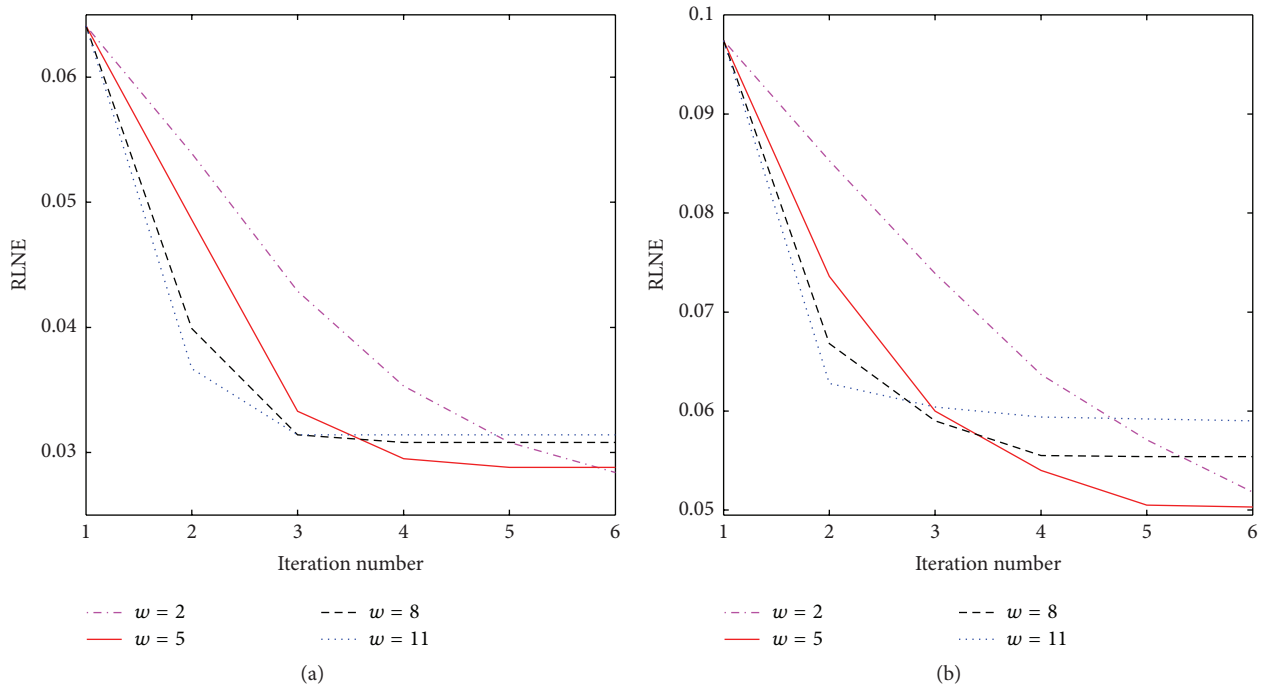

FIGURE 5: The plots of RLNEs versus iteration number for different choices of w on Brain-1 (a) and Brain-2 (b).

12 radial lines, Brain-1 and Brain-2 under 30% sampling ratio, using two-direction ICD. Table 6 presents the RLNE comparison results, which reveals that the two-direction ICD also performs well. However, compared with four-direction ICD, we see that the local features of the images can be better preserved by adding diagonal components.

4.5. Computational Complexity. The proposed method and the methods used for comparison except NLTV were implemented using a nonlinear conjugate gradient method with backtracking line-search. In a MatLab implementation, it took 10 s in average for the proposed ICD-TR and ICD-TH containing only one regularization term, while it took

TABLE 5: RLNEs of the reconstructed images using radial and Cartesian k-space sampling.

Test image	Method	Radial	Cartesian
Brain-1	SparseMRI	0.1256	0.0833
	NLTV	0.0882	0.0568
	ISD-TV	0.0963	0.0510
	SDBS-TV	0.0771	0.0424
	ICD	0.0669	0.0335
Brain-2	SparseMRI	0.1434	0.0998
	NLTV	0.1249	0.0843
	ISD-TV	0.1138	0.0806
	SDBS-TV	0.1082	0.0712
	ICD	0.0937	0.0608

TABLE 6: Comparison results of RLNEs using two- and four-direction ICD.

Test image	Directional components	Method		
		ICD-TR	ICD-TH	ICD-WT
Shepp Logan	Two	0.0254	0.0205	0.0169
	Four	0.0098	0.0042	0.0037
Brain-1	Two	0.0388	0.0354	0.0321
	Four	0.0309	0.0267	0.0257
Brain-2	Two	0.0661	0.0627	0.0607
	Four	0.0550	0.0514	0.0491

more than 50 s for other methods with an additional wavelet penalty term. However, we expect substantial reduction in the reconstruction time with code optimization and graphics processing unit implementation.

5. Conclusion

In this work, we have presented a cosparse analysis model based approach to reconstruct images from highly undersampled data using cosupport constraints. We have demonstrated that the analysis model prior can significantly improve the sparse sampling based image reconstruction. An effective iterative algorithm proceeded by alternating between image reconstruction and cosupport detection. The performance of the proposed method was evaluated through simulations on synthetic and practical MR images. The results indicate that the proposed method yields considerably better performance than methods without using prior knowledge and synthesis model based methods imposing support constraints in terms of both reconstruction accuracy and subjective visual quality.

Appendix

Proof of Proposition 3

In this part, we first prove the following two lemmas that are the key elements for the proof of Proposition 3. Note that our goal is to quantify the upper bound:

$$\max_{\Lambda} \dim\left(\mathscr{W}_{\Lambda}\right) = \max_{\Lambda}\left\{|V| - |V\left(\Lambda\right)| + J\left(\Lambda\right)\right\}$$
$$= |V| - \min_{\Lambda}\left\{|V\left(\Lambda\right)| - J\left(\Lambda\right)\right\}. \tag{A.1}$$

To simplify the problem, we look at an equivalent quantity:

$$\alpha\left(\Lambda\right) \coloneqq \min_{\Lambda}\left\{|V\left(\Lambda\right)| - J\left(\Lambda\right)\right\}. \tag{A.2}$$

Lemma A.1. *Given the known part of the cosupport Λ_0 with J connected components, then for a fixed cosparsity level l, the value $\alpha(\Lambda) = \min_{\Lambda}\{|V(\Lambda)| - J(\Lambda)\}$ is achieved when $J(\Lambda) = J$.*

Proof. If $J(\Lambda) > J$, there exist extra connected components out of $V(\Lambda_0)$. Without loss of generality, we only consider the case where $J(\Lambda \setminus \Lambda_0) = 1$. Let $\Lambda^1 \subset \Lambda \setminus \Lambda_0$ correspond to the extra component. Thus, we have $J(\Lambda) = J + 1$. Notice that a subgraph can be shifted horizontally or vertically unless it has vertices on all four boundaries of V. Since Λ^1 and Λ_0 are disconnected, we can shift Λ^1 towards Λ_0 until they first met. Suppose t and s are the numbers of the vertices and the edges coincided. Clearly, $s \leq t - 1$. The resulting subgraph Λ^2 has $|V(\Lambda)| - t$ vertices, $|\Lambda| - s$ edges, and J connected components. Then, we add s additional edges that are connected to Λ^2, thereby yielding Λ^3. From this we can infer that $|\Lambda^3| = |\Lambda| = l$, $J(\Lambda^3) = J$, and

$$\left|V\left(\Lambda^3\right)\right| \leq \left|V\left(\Lambda^2\right)\right| + s = |V\left(\Lambda\right)| - t + s. \tag{A.3}$$

We further have that

$$\left|V\left(\Lambda^3\right)\right| - J\left(\Lambda^3\right) \leq |V\left(\Lambda\right)| - t + s - J\left(\Lambda\right) + 1$$
$$\leq |V\left(\Lambda\right)| - J\left(\Lambda\right). \tag{A.4}$$

Consequently, we obtain Λ^3 from Λ that satisfies $|\Lambda^3| = l$ and $J(\Lambda^3) = J$. However, a smaller $\alpha(\Lambda)$ is achieved, which

completes the proof. One can consider cases in which $J(\Lambda \setminus \Lambda_0) > 1$ in the same vein. □

Lemma A.1 implies that quantifying the upper bound in (A.1) is equivalent to quantifying the value $\min_{|\Lambda|=l}|V(\Lambda)|$. In the next lemma, we provide a lower bound for $|V(\Lambda)|$ and prove that the value is achieved when all the J connected components share the same cosparsity.

Lemma A.2. *For a fixed l, the value*

$$|V(\Lambda)| \geq \frac{l}{2} + \frac{J}{2} + \left(\frac{Jl}{2} + \frac{J^2}{4} \right)^{1/2}. \tag{A.5}$$

Proof. We first denote the edge sets of the J connected components by $\Lambda_i \subset \Lambda$ with $|\Lambda_i| = l_i$ ($i \in I = \{1, \ldots, J\}$). Clearly, we have $\sum_i l_i = l$. Then, we enlarge each edge set Λ_i as

$$\overline{\Lambda}_i = \left\{ e_\downarrow(v), e_\rightarrow(v) : v \in V(\Lambda_i) \right\}, \tag{A.6}$$

where $e_\downarrow(v)$ (resp., $e_\rightarrow(v)$) denotes the edge extending downwards (resp., rightwards) from the vertex v. For any $i \in I$, we have $|\overline{\Lambda}_i| = 2|V(\Lambda_i)|$. According to [17, Lemma 15], the value $\min_{|\Lambda_i|=l_i}|V(\Lambda_i)|$ occurs when $V(\Lambda_i)$ is square, and the optimal width of $V(\Lambda_i)$ is given as

$$w_i = \frac{1}{2} + \left(\frac{l_i}{2} + \frac{1}{4} \right)^{1/2}. \tag{A.7}$$

Then, we have

$$2|V(\Lambda_i)| = |\overline{\Lambda}_i| \geq l_i + 2w_i. \tag{A.8}$$

Summing the above over all $i \in I$ yields

$$2|V(\Lambda)| = \sum_{i \in I} 2|V(\Lambda_i)|$$
$$\geq \sum_{i \in I} l_i + 2\sum_{i \in I} \left(\frac{1}{2} + \left(\frac{l_i}{2} + \frac{1}{4} \right)^{1/2} \right) \tag{A.9}$$
$$= l + J + \sum_{i \in I} (2l_i + 1)^{1/2}.$$

For any $j \in I$,

$$l_j = l - \sum_{i \in I \setminus j} l_i. \tag{A.10}$$

Substituting (A.10) into (A.9),

$$2|V(\Lambda)| \geq l + J + \sum_{i \in I \setminus j} (2l_i + 1)^{1/2} + \left(2l_j + 1 \right)^{1/2}$$
$$= l + J + \sum_{i \in I \setminus j} (2l_i + 1)^{1/2} \tag{A.11}$$
$$+ \left(2\left(l - \sum_{i \in I \setminus j} l_i \right) + 1 \right)^{1/2}.$$

Then for any $i \neq j$, the minimum of the right hand side of (A.11) is achieved at $l_i = l - \sum_{i \in I \setminus j} l_i = l_j$. This reveals that all the J connected components share the same cosparsity. Therefore, by denoting $l_i = l/J$ ($i \in I$) and substituting into (A.9), we conclude that

$$|V(\Lambda)| \geq \frac{l}{2} + \frac{J}{2} + \left(\frac{Jl}{2} + \frac{J^2}{4} \right)^{1/2} \tag{A.12}$$

as required. □

Proof of Proposition 3. The proof of Proposition 3 is trivial by applying Lemmas A.1 and A.2 to (A.1):

$$\max_\Lambda \dim(\mathscr{W}_\Lambda) = |V| - \min_\Lambda \{|V(\Lambda)| - J(\Lambda)\}$$
$$= n - \min_\Lambda |V(\Lambda)| + J$$
$$= n - \frac{l}{2} - \frac{J}{2} - \left(\frac{Jl}{2} + \frac{J^2}{4} \right)^{1/2} + J \tag{A.13}$$
$$\leq n - \frac{l}{2} - \sqrt{\frac{Jl}{2}} + \frac{J}{2},$$

which leads to the minimum measurement number:

$$m \geq 2n - l - \sqrt{2Jl} + J \geq 2\max_\Lambda \dim(\mathscr{W}_\Lambda). \tag{A.14}$$

The proof is then completed. □

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[4] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[5] M. Elad, *Sparse and Redundant Representations—from Theory to Applications in Signal and Image Processing*, Springer, New York, NY, USA, 2010.

[6] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, New York, NY, USA, 3rd edition, 2008.

[7] J. L. Starck, F. Murtagh, E. J. Candès, and D. L. Donoho, *Sparse Image and Signal Processing—Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press, Cambridge, UK, 2010.

[8] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[9] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis MCA," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[10] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[11] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[12] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[13] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[14] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $l1$ minimization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, 2003.

[15] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5030–5043, 2010.

[16] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.

[17] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[18] S. Vaiter, G. Peyré, C. Dossal, and J. Fadili, "Robust sparse analysis regularization," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2001–2016, 2013.

[19] R. Giryes, Y. Plan, and R. Vershynin, "On the effective measure of dimension in the analysis cosparse model," *IEEE Transactions on Information Theory*, vol. 61, no. 10, pp. 5745–5753, 2015.

[20] R. Giryes and D. Needell, "Near oracle performance and block analysis of signal space greedy methods," *Journal of Approximation Theory*, vol. 194, pp. 157–174, 2015.

[21] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.

[22] I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors," in *Wavelets XIII*, vol. 7446 of *Proceedings of SPIE*, San Diego, Calif, USA, August 2009.

[23] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

[24] R. Rubinstein, T. Peleg, and M. Elad, "Analysis K-SVD: a dictionary-learning algorithm for the analysis sparse model," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 661–677, 2013.

[25] M. Genzel and G. Kutyniok, "Asymptotic analysis of inpainting via universal shearlet systems," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2301–2339, 2014.

[26] Y. P. Liu, M. De Vos, and S. Van Huffel, "Compressed sensing of multichannel EEG signals: the simultaneous cosparsity and low-rank optimization," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 2055–2061, 2015.

[27] J. S. Jørgensen, C. Kruschel, and D. A. Lorenz, "Testable uniqueness conditions for empirical assessment of undersampling levels in total variation-regularized X-ray CT," *Inverse Problems in Science and Engineering*, vol. 23, no. 8, pp. 1283–1305, 2015.

[28] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 462–491, 2010.

[29] N. Vaswani and W. Lu, "Modified-CS: modifying compressive sensing for problems with partially known support," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '09)*, pp. 488–492, Seoul, Republic of Korea, June 2009.

[30] D. Liang, E. V. R. DiBella, R.-R. Chen, and L. Ying, "K-t ISD: dynamic cardiac MR imaging using compressed sensing with iterative support detection," *Magnetic Resonance in Medicine*, vol. 68, no. 1, pp. 41–53, 2012.

[31] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2334–2345, 2008.

[32] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.

[33] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[34] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty, "An efficient algorithm for compressed MR imaging using total variation and wavelets," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.

[35] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[36] J. Yang, Y. Zhang, and W. Yin, "A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 288–297, 2010.

[37] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[38] M. Kabanava and H. Rauhut, "Cosparsity in compressed sensing," in *Compressed Sensing and Its Applications*, H. Boche, R. Calderbank, G. Kutyniok, and J. Vybíral, Eds., pp. 315–339, Springer, New York, NY, USA, 2015.

[39] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[40] J. Huang and F. Yang, "Compressed magnetic resonance imaging based on wavelet sparsity and nonlocal total variation," in *Proceedings of the 9th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '12)*, pp. 968–971, Barcelona, Spain, May 2012.

[41] Y. Han, H. Du, W. Mei, and L. Fang, "MR image reconstruction with block sparsity and iterative support detection," *Magnetic Resonance Imaging*, vol. 33, no. 5, pp. 624–634, 2015.

[42] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[44] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, 1993.

[45] S. Ramani, T. Blu, and M. Unser, "Monte-carlo sure: a black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1540–1554, 2008.

[46] S. Ramani, Z. Liu, J. Rosen, J.-F. Nielsen, and J. A. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3659–3672, 2012.